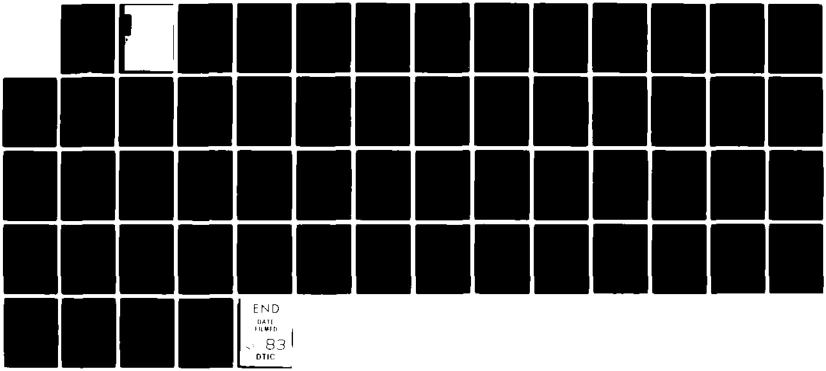
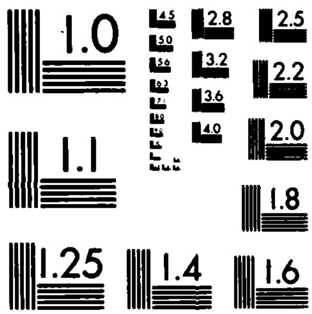
AD-A123 985 CODING FOR CORRELATED SOURCES WITH UNKNOWN PARAMETERS (U) ILLINOIS UNIV AT URBANA COORDINATED SCIENCE LAB M S WALLACE MAY 80 R-884 N00014-79-C-0424 1/1

UNCLASSIFIED F/G 9/4 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A123985	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CODING FOR CORRELATED SOURCES WITH UNKNOWN PARAMETERS		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER R-884; UILU-ENG 80-2216
7. AUTHOR(s) Mark Stanley Wallace		8. CONTRACT OR GRANT NUMBER(s) ENG75-20864 and N00014-79-C-0424
9. PERFORMING ORGANIZATION NAME AND ADDRESS Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Joint Services Electronics Program		12. REPORT DATE May 1980
		13. NUMBER OF PAGES 55
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

source coding, correlated discrete memoryless sources, robust coding, side information

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A number of problems in source coding deal with a pair of correlated discrete memoryless sources and two separate non-cooperating encoders, such as the Slepian-Wolf problem, the side-information problem, and the Wyner-Ziv problem. In these problems it is desired to determine rate pairs which allow the outputs of each of two sources to be reproduced at the decoder with some specified distortion. If the joint distribution of the outputs of the two sources is completely known to both encoders, then solutions to these problems are available in the literature. Here the situation in which the joint distribution is not completely known is considered.

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CODING FOR CORRELATED SOURCES WITH UNKNOWN PARAMETERS

by

Mark Stanley Wallace

This work was supported in part by the National Science Foundation under Grant ENG75-20864 and in part by the Joint Services Electronics Program (U.S. Army, U.S. Navy and U.S. Air Force) under Contract N00014-79-C-0424.

CODING FOR CORRELATED SOURCES WITH UNKNOWN PARAMETERS

BY

MARK STANLEY WALLACE

B. Eng., McGill University, 1978

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1980

Thesis Advisor: Professor Michael B. Pursley

Urbana, Illinois

CODING FOR CORRELATED SOURCES WITH UNKNOWN PARAMETERS

by

MARK STANLEY WALLACE
Coordinated Science Laboratory and
Department of Electrical Engineering
University of Illinois at Urbana-Champaign

ABSTRACT

A number of problems in source coding deal with a pair of correlated discrete memoryless sources and two separate non-cooperating encoders, such as the Slepian-Wolf problem, the side-information problem, and the Wyner-Ziv problem. In these problems it is desired to determine rate pairs which allow the outputs of each of two sources to be reproduced at the decoder with some specified distortion. If the joint distribution of the outputs of the two sources is completely known to both encoders, then solutions to these problems are available in the literature. Here the situation in which the joint distribution is not completely known is considered.

The work was supported by the National Science Foundation under Grant ENG75-20864 and by the Joint Services Electronics Program under Contract N00014-79-C-0424.

ACKNOWLEDGEMENTS

The author is grateful to Professor M. B. Pursley for his advice and assistance throughout the preparation of this thesis.

Thanks are also due to Mrs. Phyllis Young and Ms. Christine Cox for their expert typing.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

A number of problems in source coding deal with a pair of correlated discrete memoryless sources and two separate non-cooperating encoders as in Fig. 1. Among these are the Slepian-Wolf problem [1], the side-information problem [2], [3], and the Wyner-Ziv problem [4]. The general goal of these problems is to determine rate pairs (R_1,R_2) which allow the outputs of each of two sources to be reproduced at the decoder with some specified distortion. The rate R_i is the rate of transmission of information from the i-th source encoder to the decoder. If the joint distribution* p, where $p(u,v) = P(X^{(1)} = u, X^{(2)} = v)$, is completely known to both encoders, then solutions to these problems are known. Here we consider these problems assuming only that the joint distribution is known to be in some class Λ.

For problems with only a single source, universal codes are source codes which achieve some performance measure (e.g., the entropy or the rate-distortion function) asymptotically for all distributions in some class [12]. For the Slepian-Wolf and side-information problems, where decoding with arbitrarily low distortion is required, universal coding is not possible in general. This can be seen as follows. Let the marginal distribution for $X^{(i)}$ be denoted by p_i. The rate pair for any code depends only on the marginal distributions. Thus, for a given code only a single rate pair (R_1,R_2) is possible for any set of joint distributions which

*Note that since $X^{(1)}$ and $X^{(2)}$ (the random variables representing the source outputs) are discrete, all of the distributions here are probability mass functions.

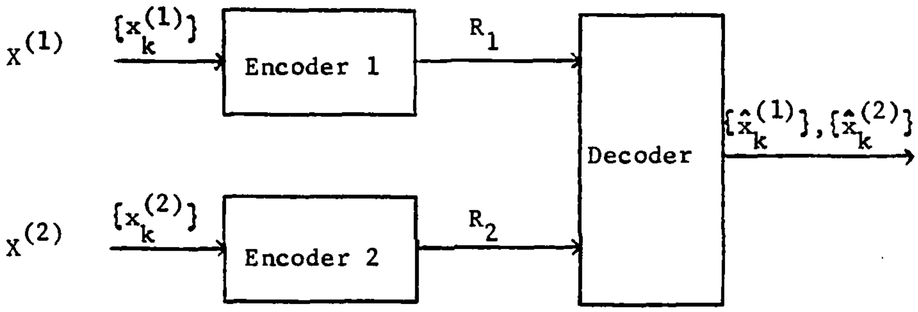


Figure 1. Encoder-decoder configuration.

have the same marginal distributions p_1 and p_2. The encoders cannot distinguish between different joint distributions which have the same marginals, since each encoder observes the output of only one of the two sources. So if the true joint distribution is π then all $p \in \Lambda$ such that $p_i = \pi_i$, $(i = 1,2)$ remain as possible distributions as far as the source encoders are concerned even if the marginals π_1 and π_2 are known exactly. However, a single rate pair (R_1,R_2) must be used for all of them. Therefore, universal coding is possible only if those $p \in \Lambda$ with the same marginals also have the same achievable rate regions.

For the Wyner-Ziv problem the same reasoning shows that a single rate pair must correspond to each set of $p \in \Lambda$ with a given pair (p_1,p_2) of marginal distributions. However, a code may still be universal if it achieves the optimum distortion for each such p. Universal fixed-rate codes for single sources do exist if some positive distortion is allowed [13].

If a code for a class of sources achieves a point (R,D) on the rate distortion curve for one source in the class and yields rate not greater than R and distortion not greater than D for all other sources in the class, it will be called a <u>robust</u> code [14]. Here we show that robust codes do exist for the Slepian-Wolf and side-information problems. For the Wyner-Ziv problem an example is given showing that robust coding is not possible in general (so universal coding is not possible either).

Suppose that the true distribution is given by p. Then the i-th encoder can only reduce Λ to a subset $\Lambda_i(p_i) \triangleq \{\pi \in \Lambda: \pi_i = p_i\}$, which is the set of distributions with the same i-th marginal as p. In the cases where robust coding is possible, optimal performance will be achieved by having the i-th encoder estimate p_i and then do robust coding for $\Lambda_i(p_i)$. The sets $\Lambda_1(p_1)$ and $\Lambda_2(p_2)$ may be different, so the choice of rates for this coding is not simple in general. In the side-information problem only $\Lambda_1(p_1)$ is used, and hence a coding technique which gives optimal performance is easily obtained. This is not the case with the Slepian-Wolf problem. A non-computable characterization of the set of achievable rate pairs for this problem is given in [7] and can be described as follows. Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be the spaces of possible marginals for $X^{(1)}$ and $X^{(2)}$ respectively. A rate pair (R_1,R_2) is achievable for a given (π_1,π_2) if there exist functions $f_1: \mathbb{P}_1 \to [0,\infty)$ and $f_2: \mathbb{P}_2 \to [0,\infty)$ such that $R_i = f_i(\pi_i)$ and also such that for all $p \in \Lambda$ the rate pair $(f_1(p_1), f_2(p_2))$ is in the Slepian-Wolf region for the pair of sources with joint distribution p. In Chapter 6 an upper bound to the set of achievable rate pairs is given, and a number of such functions f_1 and f_2 which yield sets of achievable rate pairs are considered.

By way of introduction a special case of the Slepian-Wolf problem (which is also a special case of the side-information problem) is considered in Chapter 3. The robust coding result for the side-information case is derived in Chapter 4. Chapter 5 concerns the Wyner-Ziv problem. Here an example is given which proves that robust coding is not possible in general, and a special case is presented where robust coding is possible.

CHAPTER 2

PRELIMINARIES

An encoder may make an estimate of its marginal distribution (which is with high probability within a prescribed accuracy) by observing the relative frequency of source letter outputs during some initial estimation time. Similarly if the encoders send information at rate R_i equal to the entropy of the i-th source $(i = 1,2)$ for some initial time, the decoder can estimate the joint distribution. The arbitrarily small uncertainties in these estimates have no effect on the achievable rate regions if the class Λ is assumed to be closed. Hence, in the body of the thesis it is assumed that the marginal distribution for the i-th source is known to i-th encoder $(i = 1,2)$ and the joint distribution is known to the decoder. In addition it is assumed that for some $\alpha > 0$, $\pi(u,v) \geq \alpha$ for all $(u,v) \in \mathcal{X}^{(1)} \times \mathcal{X}^{(2)}$ and all $\pi \in \Lambda$. The results needed to prove the coding theorems without these assumptions are derived in Appendix A.

The alphabet for the i-th source is denoted by $\mathcal{X}_i$ which is a finite set with $|\mathcal{X}_i|$ elements. The output of the i-th source at time k is a random variable denoted $X_k^{(i)}$ with marginal distribution p_i; i.e., $P\{X_k^{(i)} = u\} = p_i(u)$ for $u \in \mathcal{X}_i$. The random variables $(X_k^{(1)}, X_k^{(2)})$ have joint distribution p and are independent of $(X_j^{(1)}, X_j^{(2)})$ for $j \neq k$. Define $\underline{X}^{(i)}$ to be a sequence of n outputs from the i-th source $\{X_1^{(i)}, \ldots, X_n^{(i)}\}$, and let $\mathcal{X}_i^n$ be the set of all n-sequences whose components are elements of $\mathcal{X}_i$. A blocklength n code for the i-th source is a function $g_i : \mathcal{X}_i^n \to \{1,2,\ldots,\|g_i\|\}$. The rate of this code is given by

$$R_i \triangleq \frac{1}{n} \log \|g_i\| .$$

(Base two logarithms are used throughout.) The decoder is defined by a function

$$f : \{1,\ldots,\|g_1\|\} \times \{1,\ldots,\|g_2\|\} \to \hat{\mathcal{X}}_1^n \times \hat{\mathcal{X}}_2^n$$

where $\hat{\mathcal{X}}_i$ is the reproduction alphabet for the i-th source. Thus the decoder output which corresponds to the source output vectors $(\underline{X}^{(1)},\underline{X}^{(2)})$ is

$$(\underline{\hat{X}}^{(1)},\underline{\hat{X}}^{(2)}) \triangleq f[g_1(\underline{X}^{(1)}),g_2(\underline{X}^{(2)})] \ .$$

The distortion achieved by a code is

$$E\{d(\underline{X}^{(i)},\underline{\hat{X}}^{(i)})\} \triangleq E\{n^{-1}\sum_{j=1}^{n} d_i(X_j^{(i)},\hat{X}_j^{(i)})\} \ ,$$

the expected per letter distortion between $\underline{X}^{(i)}$ and $\underline{\hat{X}}^{(i)}$ computed using a single letter distortion measure $d_i(\cdot,\cdot)$ defined on $\mathcal{X}_i \times \hat{\mathcal{X}}_i$. A rate pair (R_1,R_2) will be called achievable if for all positive ϵ and δ, codes of rates $R_i' < R_i + \epsilon$ exist for which $E\{d(\underline{X}^{(i)},\underline{\hat{X}}^{(i)})\} < D_i + \delta$ where D_i is the specified distortion level for the i-th source. The set of <u>all</u> achievable rate pairs will be called the achievable rate region.

Let $\underline{x}^{(i)}$ denote the n-vector $\{x_1^{(i)},\ldots,x_n^{(i)}\}$ where $x_j^{(i)} \in \mathcal{X}_i$ for $1 \le j \le n$, and define $N(u;\underline{x}^{(i)})$ to be the number of occurrances of the letter u in the sequence $\underline{x}^{(i)}$. A sequence $\underline{x}^{(i)}$ is called δ-typical if

$$|n^{-1}N(u;\underline{x}^{(i)})-p_i(u)| < \delta|\mathcal{X}_i|^{-1} \qquad (1)$$

for all $u \in \mathcal{X}_i$, $\delta > 0$. The set of δ-typical $\underline{x}^{(i)}$ sequences will be denoted $T_n(\delta,p_i)$. The set $T_n(\delta,p_i)$ has the following properties which are proved in [5] and [11].

<u>Property 1</u>. For any $\delta > 0$, $P[\underline{X}^{(i)} \in T_n(\delta,p_i)] \to 1$ as $n \to \infty$ uniformly on Λ.

<u>Property 2</u>. $\exp_2[n(\mathcal{H}(p_i)-c)] < |T_n(\delta p_i)| < \exp_2[n(\mathcal{H}(p_i)+c)]$

where

$$c \triangleq -\delta \log \alpha^{-1} \geq -\delta\, |\mathcal{X}_i| \sum_{u \in \mathcal{X}_i} \log p_i(u)$$

and

$$\mathcal{H}(p_i) = -\sum_{u \in \mathcal{X}_i} p_i(u) \log p_i(u) \ .$$

(Here $\exp_2(a) = 2^a$ for any real number a.)

Similarly define $N(u,v;\underline{x}^{(1)},\underline{x}^{(2)})$ to be the number of indices j such that $(u,v) = (x_j^{(1)},x_j^{(2)})$, $1 \leq j \leq n$, the $x_j^{(i)}$ being the components of $\underline{x}^{(i)}$. Then call $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$ jointly δ-typical if

$$|n^{-1}N(u,v;\underline{x}^{(1)},\underline{x}^{(2)}) - p(u,v)| < \delta |\mathcal{X}_1|^{-1} |\mathcal{X}_2|^{-1} \qquad (2)$$

for all $(u,v) \in \mathcal{X}_1 \times \mathcal{X}_2$ where p is the joint distribution of $X^{(1)}$ and $X^{(2)}$. The set of jointly δ-typical $(\underline{x}^{(1)},\underline{x}^{(2)})$ will be denoted $J_n(\delta,p)$. Note that

$$(\underline{x}^{(1)},\underline{x}^{(2)}) \in J_n(\delta,p) \Rightarrow \underline{x}^{(i)} \in T_n(\delta,p_i) \ .$$

If the pair of random vectors $(\underline{X}^{(1)},\underline{X}^{(2)})$ is considered a single vector of independent random variables $\{(X_1^{(1)},X_1^{(2)}),\ldots,(X_n^{(1)},X_n^{(2)})\}$ each with distribution p, then the defining inequality (1) for δ-typical is equivalent to inequality (2). Thus directly from the properties of $T_n(\delta,p_i)$ we have

$$P[(\underline{X}^{(1)},\underline{X}^{(2)}) \in J_n(\delta,p)] \to 1 \quad \text{as } n \to \infty$$

and

$$\exp_2[n(\mathcal{H}(p)-c)] < |J_n(\delta,p)| < \exp_2[n(\mathcal{H}(p)+c)]$$

where $\mathcal{H}(p)$ is the joint entropy of $X^{(1)}$ and $X^{(2)}$.

For the Slepian-Wolf and side information problems we define a robust code as follows. Let $\mathfrak{R}(\pi)$ denote the set of boundary points of the achievable rate region for the case in which the joint distribution π is known to both encoders and let $R_i(n,p)$ be the rate of a code of blocklength n for the i-th encoder when the source distribution is p. A sequence of codes of increasing blocklength n for a class of sources Λ is robust if the codes achieve zero distortion uniformly as $n \to \infty$ for all sources in Λ and if for some $\pi \in \Lambda$ the rate pairs $(R_1(n,p),R_2(n,p))$ (which may be a function of p the true source distribution) satisfy

$$R_i(n,p) < R_i^* + \epsilon_N, \quad i = 1,2,$$

for all $p \in \Lambda$ and $n \geq N$, where $\epsilon_N \to 0$ as $N \to \infty$ and is independent of p and $(R_1^*,R_2^*) \in \mathfrak{R}(\pi)$. So a sequence of codes is robust if it achieves asymptotically zero distortion for all sources in the class.

To modify this definition for the Wyner-Ziv problem let $R_\pi^*(D)$ be the Wyner-Ziv rate distortion function for source π. Then call a sequence of blocklength n codes robust if the rate $R(n,\pi)$ and distortion $D(n,\pi)$ converge to a point on $R_\pi^*(D)$ for some source $\pi \in \Lambda$ and if $R_1(n,\hat{\pi}) \leq R_1(n,\pi)$ and $D(n,\hat{\pi}) \leq D(n,\pi)$ for all $\hat{\pi} \in \Lambda$.

CHAPTER 3

ROBUST CODING FOR THE CORNER POINT OF THE SLEPIAN-WOLF REGION

3.1 Statement of Problem and Preliminary Result

The Slepian-Wolf problem is to determine the set of rate pairs (R_1,R_2) which allow the outputs of both sources to be reproduced at the decoder with arbitrarily low distortion. In this problem $\hat{\mathcal{X}}_i = \mathcal{X}_i$, and the distortion measure is the Hamming distortion measure (i.e., $d_i(u,u) = 0$, $d_i(u,v) = 1$ if $u \neq v$), so arbitrarily low probability of decoding error corresponds to arbitrarily low distortion. Here we derive a robust coding result for the special case where $R_2 > \mathcal{H}(p_2)$. Since $R_2 > \mathcal{H}(p_2)$ the decoder may be assumed to know $\{X_k^{(2)}\}$ exactly, and the problem becomes the determination of the set of rates R_1 which allow $\{X_k^{(1)}\}$ to be recovered with arbitrarily low distortion. This is the corner point of the Slepian-Wolf region (see Fig. 2). Notice that an increase in R_2 above $\mathcal{H}(p_2)$ does not permit a decrease in R_1.

The joint distribution is known to be in some class Λ. In addition the encoder knows the marginal p_1, and the decoder knows the true joint distribution p. However, the encoder does not know the joint distribution p nor the marginal p_2. In this case R_1 is achievable if and only if

$$R_1 \geq \sup\{h_1(\pi) \; : \; \pi \in \Lambda_1(p_1)\} \tag{3}$$

where $\Lambda_1(p_1) \triangleq \{\pi \in \Lambda \; : \; \pi_1 = p_1\}$ as before, and $h_1(\pi)$ is the average conditional entropy given by

$$h_1(\pi) \triangleq -\sum_u \sum_v \pi(u,v) \log \frac{\pi(u,v)}{\sum_u \pi(u,v)} \quad . \tag{4}$$

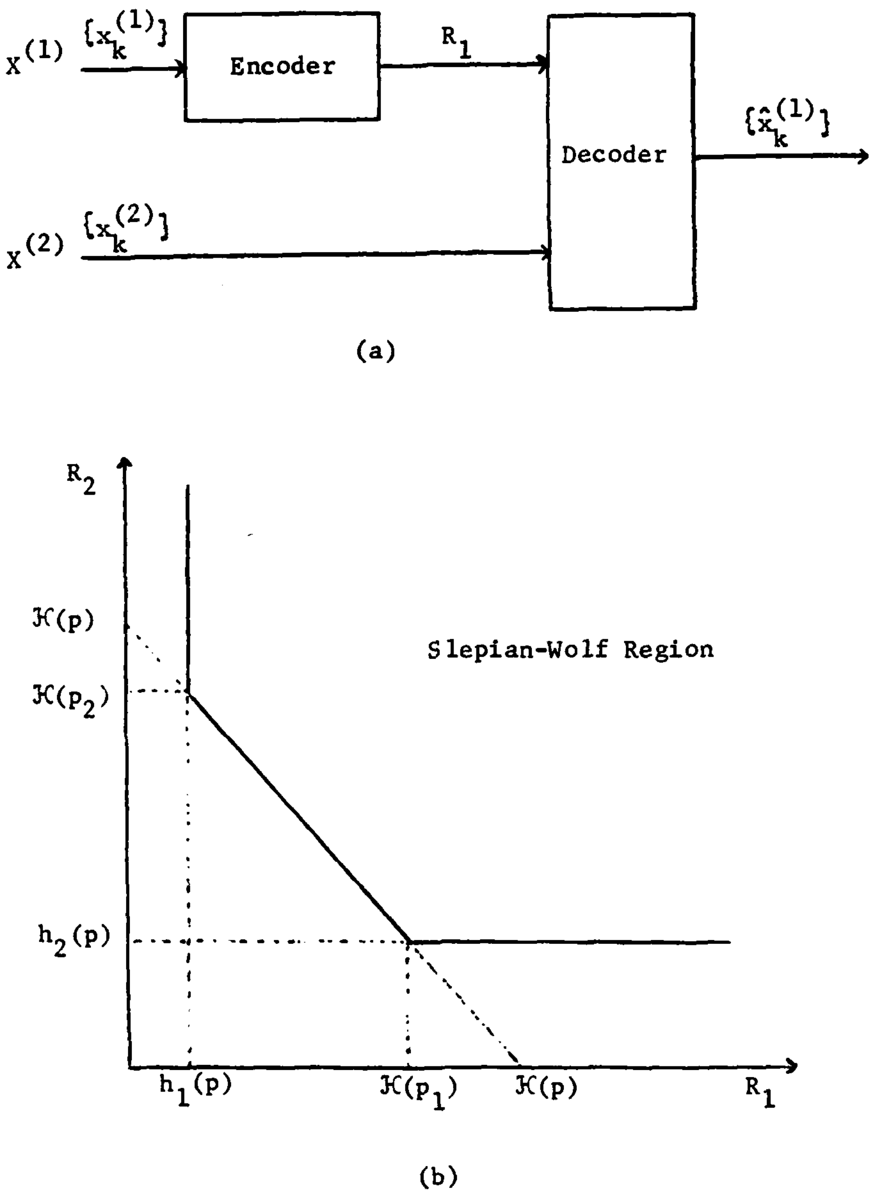


Figure 2. Slepian-Wolf problem. (a) Encoder-decoder configuration for corner point. (b) Slepian-Wolf rate region for a source p.

Notice that for any distribution p on $\mathcal{X}_1 \times \mathcal{X}_2$,

$$h_1(p) = \mathcal{H}(p) - \sum_u \sum_v p(u,v) \log p_2(v)$$

$$= \mathcal{H}(p) - \mathcal{H}(p_2) \ . \tag{5}$$

3.2 Positive Coding Theorem

Define

$$\bar{h}_1(p_1) \triangleq \sup\{h_1(\pi) : \pi \in \Lambda_1(p_1)\} \tag{6}$$

The proof of the positive coding theorem follows directly from that of Berger [5, Theorem 3.2], simply substituting $\bar{h}_1(p_1)$ for $h_1(\pi)$ and fixing $R_2 > \mathcal{H}(p_2)$. This proof with the necessary modifications is as follows.

Form $N \triangleq \exp_2[n(\bar{h}_1(p_1) + 2\gamma)]$ sets of $\underline{x}^{(1)}$ sequences, say $S_1, \ldots, S_N$, by selecting sequences independently from $T_n(\delta, p_1)$ according to a uniform distribution. The selections are made with replacement so that every $\underline{x}^{(1)}$ sequence is equally likely to be selected each time. This is done until each set S_i contains

$$|S| \triangleq \exp_2[n(\mathcal{H}(p_1) - \bar{h}_1(p_1) - \gamma)] \tag{7}$$

δ-typical $\underline{x}^{(1)}$ sequences.

A sequence $\underline{x}^{(1)}$ is encoded into an index $i(\underline{x}^{(1)})$ defined by

$$i(\underline{x}^{(1)}) = \begin{cases} j & \text{if } j = \min\{k : \underline{x}^{(1)} \in S_k\} \\ \\ 0 & \text{if } \underline{x}^{(1)} \notin S_k,\ k = j, \ldots, N \ . \end{cases} \tag{8}$$

The index $i(\underline{x}^{(1)})$ is then sent to the decoder, requiring a rate $R_1 = \bar{h}_1(p_1) + 2\gamma$. An error will occur if $i(\underline{x}^{(1)}) = 0$. Since

$$P\{\underline{X}^{(1)} \notin T_n(\delta,p_1)\} \to 0 \qquad \text{as } n \to \infty ,$$

we will know

$$P\{i(\underline{X}^{(1)}) = 0\} \to 0 \qquad \text{as } n \to \infty$$

if

$$Q \triangleq P\{i(\underline{X}^{(1)}) = 0 \mid \underline{X}^{(1)} \in T_n(\delta,p_1)\} \to 0 \qquad \text{as } n \to \infty.$$

Now

$$\begin{aligned} Q &= \prod_{j=1}^{N} P\{\underline{X}^{(1)} \notin S_j \mid \underline{X}^{(1)} \in T_n(\delta,p_1)\} \\ &= \prod_{j=1}^{N} \prod_{\underline{u}\in S_j} P\{\underline{X}^{(1)} \neq \underline{u} \mid X^{(1)} \in T_n(\delta,p_1)\} \\ &= P\{\underline{X}^{(1)} \neq \underline{u} \mid \underline{X}^{(1)} \in T_n(\delta,p_1)\}^{N|S|} \\ &= \{1 - |T_n(\delta,p_1)|\}^{N|S|} \\ &\le (1 - \exp_2[-n(\mathcal{H}(p_1) + c)])^{N|S|} \end{aligned} \tag{9}$$

since the selections were done independently using a uniform distribution. Since $\ln(1-x) \le -x$, $0 \le x < 1$, then

$$\begin{aligned} \ln Q &\le N|S| \ln\{1 - \exp_2[-n\mathcal{H}(p_1) + c)]\} \\ &\le -N|S| \exp_2[-n\mathcal{H}(p_1) + c)] \\ &= -\exp_2[n(\gamma - c)]. \end{aligned} \tag{10}$$

So $\ln Q \to -\infty$ and $P\{i(X^{(1)}) = 0\} \to 0$ as $n \to \infty$ if $\gamma > c$.

Now $P\{(\underline{X}^{(1)}, \underline{X}^{(2)}) \in J_n(\delta,p)\} \to 1$ as $n \to \infty$ so we may assume that the source outputs $(\underline{x}^{(1)}, \underline{x}^{(2)})$ are jointly δ-typical. The decoder observes $\underline{x}^{(2)}$ and knows an index j such that $\underline{x}^{(1)} \in S_j$. The decoding procedure is to search S_j for a sequence $\underline{u}$ such that $(\underline{u}, \underline{x}^{(2)}) \in J_n(\delta,p)$. As an upper bound assume an error occurs if there is some sequence $\underline{u} \in S_j$, $\underline{u} \neq \underline{x}^{(1)}$ such that $(\underline{u}, \underline{x}^{(2)}) \in J_n(\delta,p)$. Let $E(\underline{x}^{(2)})$ denote this event for a given $\underline{x}^{(2)}$. Then

$$P[E(\underline{x}^{(2)})] \leq (|S| - 1)\ P\{U \in \mathcal{U}(\underline{x}^{(2)})\}$$

where

$$\mathcal{U}(\underline{x}^{(2)}) \triangleq \{\underline{u} \in T_n(\delta,p_1) : (\underline{u}, \underline{x}^{(2)}) \in J_n(\delta,p)\}$$

and $\underline{U}$ is a random vector uniformly distributed on $T_n(\delta,p_1)$. The probability of error associated with this event is the expectation of $P[E(\underline{x}^{(2)})]$ over all δ-typical $\underline{x}^{(2)}$ sequences. Denoting this probability of error by $P[E]$ we have

$$P[E] < |S| \sum_{\underline{x}^{(2)}} P\{\underline{U} \in \mathcal{U}(\underline{x}^{(2)})\} P\{\underline{X}^{(2)} = \underline{x}^{(2)}\}$$

$$\leq |S| \exp[-n(\mathcal{H}(p_2) - c)] \sum_{\underline{x}^{(2)}} P\{\underline{U} \in \mathcal{U}(\underline{x}^{(2)}\}.$$

$$\leq |S| \exp[-n(\mathcal{H}(p_2) - c)] \sum_{\underline{x}^{(2)}} |\mathcal{U}(\underline{x}^{(2)})|\,|T_n(\delta,p_1)|^{-1} \tag{11}$$

But

$$\sum_{\underline{x}^{(2)}} |\mathcal{U}(\underline{x}^{(2)})| = |J_n(\delta,p)|$$

so

$$P[E] \le |S| \exp[-n(\mathcal{H}(p_2) - c)] \, |J_n(\delta,p)| \, |T_n(\delta,p_1)|^{-1}$$

$$< |S| \exp_2[n(\mathcal{H}(p) - \mathcal{H}(p_1) - \mathcal{H}(p_2) + 3c)] \qquad (12)$$

using the bounds on cardinality from Property 2 of the sets of typical sequences. The constant c is defined in the Preliminaries. Since

$$|S| = \exp_2[n(\mathcal{H}(p_1) - \bar{h}_1(p_1) - \gamma)], \qquad (13)$$

then it follows from (5), (12), and (13) that

$$P[E] < \exp_2[n(h_1(p) - \bar{h}_1(p_1) + 3c - \gamma)]. \qquad (14)$$

Since (6) implies $\bar{h}_1(p_1) \ge h_1(p)$, it follows from (14) that $P[E] \to 0$ as $n \to \infty$ if

$$\gamma > 3c \; . \qquad (15)$$

The constant c may be made arbitrarily small and does not depend on the distribution p, so any $R_1 > \bar{h}_1(p_1)$ is achievable.

3.3 Converse to Coding Theorem

The rate of any code for source 1 can be expressed as

$$R_1 = R_1(p_1) \triangleq \sum_{\underline{u}} P\{\underline{X}^{(1)} = \underline{u}\}\ell(\underline{u}) = n^{-1} \sum_{\underline{u}} [\prod_{i=1}^{n} p_1(u_i)]\ell(\underline{u}) \qquad (16)$$

where $\ell(\underline{u})$ is the length (i.e., number of bits or symbols) of the codeword for $\underline{u}$. This length function depends only on the code, so for a given code the rate is a function of the marginal distribution p_1. All sources $\pi \in \Lambda_1(p_1)$ have the same marginal p_1, hence any code has a single rate $R_1 = R_1(p_1)$ for all of them. By the converse of the Slepian-Wolf theorem applied to a particular source $\pi \in \Lambda_1(p_1)$

$$R_1(p_1) \geq h_1(\pi), \tag{17}$$

and since this applies to all sources in $\Lambda_1(p_1)$,

$$R_1 = R_1(p_1) \geq \sup\{h_1(\pi) : \pi \in \Lambda_1(p_1)\} = \tilde{h}_1(p_1) \tag{18}$$

which is the desired result.

CHAPTER 4

ROBUST CODING FOR THE SIDE INFORMATION PROBLEM

4.1 Statement of Problem and Result

In this problem the output of source 1 must be reproduced with arbitrarily low distortion by the decoder. Encoder 2 sends information derived from source 2 at rate R_2 and this information is used to determine the $\{\hat{X}^{(1)}\}$ but reproduction of the output of source 2 is not required (see Fig. 3). The solution to this problem is given in terms of an auxiliary random variable Z. Let the joint distribution of $X^{(1)}$, $X^{(2)}$ and Z be

$$q(u,v,w) \triangleq P\{X^{(1)} = u, X^{(2)} = v, Z = w\}$$

and the joint distribution of $X^{(i)}$ and Z be

$$q_i(u,v) \triangleq P\{X^{(i)} = u, Z = v\}$$

for $i = 1,2$. Also, define p_z to be the marginal distribution of Z. In the case where the joint distribution p of $X^{(1)}$ and $X^{(2)}$ is known precisely by the encoders and the decoder the rate region is ([2],[3])

$$\mathfrak{R} = \{(R_1,R_2): \quad Z \quad X^{(1)} \rightarrow X^{(2)} \rightarrow Z \text{ and } R_1 \geq h_z(q_1),\ R_2 \geq \mathcal{J}(q_2)\}. \quad (19)$$

Here $X^{(1)} \rightarrow X^{(2)} \rightarrow Z$ indicates that these random variables in the order listed form a Markov chain (i.e, $X^{(1)}$ and Z are conditionally independent given $X^{(2)}$),

$$h_z(q_i) \triangleq -\sum_{u,v} q_i(u,v) \log \frac{q_i(u,v)}{\sum_u q_i(u,v)}$$

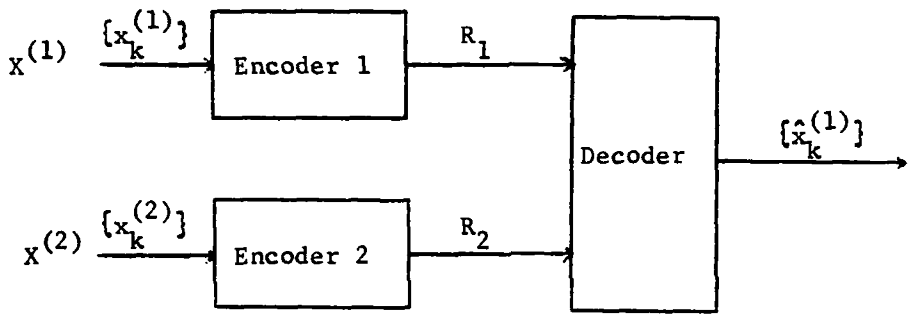


Figure 3. Configuration for the side information problem.

$$= \mathcal{H}(q_i) - \mathcal{H}(p_z) \tag{20}$$

is the conditional entropy of $X^{(i)}$ given Z, and

$$\mathcal{J}(q_2) \overset{\Delta}{=} -\sum_{u,v} q_2(u,v) \log \frac{q_2(u,v)}{p_2(u)p_z(v)}$$

$$= \mathcal{H}(p_2) - h_z(q_2)$$

is the mutual information between $X^{(2)}$ and Z. In all that follows, we assume only that $p \in \Lambda$, that the i-th encoder knows the i-th marginal p_i, and that the decoder knows p.

Since $X^{(1)} \to X^{(2)} \to Z$, then $q(x^{(1)},x^{(2)},z) = p(x^{(1)},x^{(2)})w(z|x^{(2)})$ and so the auxiliary random variable Z may be specified by a conditional distribution w where $w(u|v) = P\{Z = u|X^{(2)} = v\}$. This conditional distribution is chosen beforehand and is known to the encoders and the decoder. Notice that q_1 is obtained from p and w by

$$q_1(x^{(1)},z) = \sum_{x^{(2)}} p(x^{(1)},x^{(2)})w(z|x^{(2)}). \tag{21}$$

Under these assumptions the rate region is

$$\hat{\mathcal{R}} = \{(R_1,R_2) : \exists\, w \ni R_1 \geq \bar{h}_z(p_1),\ R_2 \geq \mathcal{J}(q_2)\} \tag{22}$$

where

$$\bar{h}_z(p_1) \overset{\Delta}{=} \sup\{h_z(q_1) : \pi \in \Lambda_1(p_1)\} \tag{23}$$

and

$$\Lambda_1(p_1) \overset{\Delta}{=} \{\pi \in \Lambda : \pi_1 = p_1\}.$$

Note that $\bar{h}_z(p_1)$ is defined for a fixed w, and so q_1 is a function of π as indicated in (21) with p replaced by π.

4.2 Positive Coding Theorem

The proof here is that of Berger ([5], Theorem 5.1) with a few modifications, as in the positive coding theorem for the Slepian-Wolf problem in section 3.2. Encoder 2 knows p_2 and w so it may encode exactly as if the distributions were known. This is done as follows. Let D be a subset of $T_n(\delta, p_z)$ of size

$$|D| \stackrel{\Delta}{=} \exp_2[n(\mathcal{J}(q_2) + f(\delta))], \tag{24}$$

for which the sequences $\underline{z}$ are chosen from $T_n(\delta, p_z)$ according to a uniform distribution. In (24), $f(\delta)$ does not depend on p since $p(u) \geq \alpha > 0$ and $f(\delta) \to 0$ as $\delta \to 0$. It can be shown (Lemma 2.1.3 of [5]) that if $\underline{x}^{(2)}$ is any δ-typical output sequence from source 2, there will be at least one $\underline{z} \in D$ such that $\underline{x}^{(2)}$ and $\underline{z}$ are jointly δ-typical w.p. $\to 1$ (w.p. $\to \alpha$ will be used to indicate with probability $\to \alpha$ as $n \to \infty$). Let D be such a set. If the source output is $\underline{x}^{(2)}$, encoder 2 simply sends the smallest index which corresponds to a $\underline{z}$ which is in D and has the property that $(\underline{x}^{(2)}, \underline{z}) \in J_n(\delta, q_2)$, and sends index 0 if there is no such $\underline{z}$. Then we have

$$R_2 = \mathcal{J}(q_2) + f(\delta) \tag{25}$$

where $f(\delta) \to 0$ as $\delta \to 0$.

Encoder 1 first determines $\bar{h}_z(p_1)$ using Λ, p_1, and w (see (23)). Next

$$N \triangleq \exp[n(\bar{h}_z(p_1) + 2\gamma)] \tag{26}$$

sets of $\underline{x}^{(1)}$ sequences, say $S_1, \ldots, S_N$, are formed by selecting sequences independently from $T_n(\delta,p_1)$ according to a uniform distribution, as in section 3.2. Here each set contains

$$|S| \triangleq \exp_2[n\mathcal{H}(p_1) - \bar{h}_z(p_1) - \gamma] \tag{27}$$

δ-typical $\underline{x}^{(1)}$ sequences. The index $i(\underline{x}^{(1)})$ is defined by

$$i(\underline{x}^{(1)}) = \begin{array}{ll} j & \text{if } j = \min\{k : \underline{x}^{(1)} \in S_k\} \\ 0 & \text{if } \underline{x}^{(1)} \notin S_k,\ k = 1, \ldots, N \end{array}$$

and this index is sent to the decoder requiring a rate $R_1 = \bar{h}_z(p_1) + 2\gamma$. With N and $|S|$ as defined in (26) and (27), the derivation of equations (9) and (10) in section 3.2 holds, showing that $P\{i(\underline{X}^{(1)}) = 0\} \to 0$ as $n \to \infty$.

Since $P\{(\underline{X}^{(1)},\underline{X}^{(2)}) \in J_n(\delta,p)\} \to 1$ as $n \to \infty$ we may assume that the source outputs $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$ are jointly typical, i.e., $(\underline{x}^{(1)},x^{(2)}) \in J_n(\delta,p)$. The decoder observes a sequence $\underline{z}$ jointly δ-typical with $\underline{x}^{(2)}$ and knows an index j such that $\underline{x}^{(1)} \in S_j$. Lemma 4.1 of [5] ("Markov Lemma") states

$$X^{(1)} \to X^{(2)} \to Z \text{ and } (\underline{X}^{(2)},\underline{z}) \in J_n(\delta,q_2) \Rightarrow (\underline{X}^{(1)},\underline{z}) \in J_n(\delta|\mathcal{X}_1|,q_1)$$

w.p. $\to$ 1. The decoding procedure is to search S_j for a sequence $\underline{u}$ such that $(\underline{u},\underline{z}) \in J_n(\delta_1,q_1)$, where $\delta_1 = \delta|\mathcal{X}_1|$. The set S_j contains such a $\underline{u}$ w.p. $\to$ 1 by the above lemma. Let $E_{\underline{z}}$ be the event that at least one of the $|S| - 1$ sequences $\underline{u} \in S_j$ other than the sequence $\underline{x}^{(1)}$ (the actual source output) is δ_1-typical with $\underline{z}$. If an error occurs then either $E_{\underline{z}}$ occurs or else there is no $\underline{u} \in S_j$ such that $(\underline{u},\underline{z}) \in J_n(\delta_1,q_1)$. The latter event has probability 0 asymptotically as we mentioned above, and the event $E_{\underline{z}}$ has probability upper bounded by

$$P(E_{\underline{z}}) \le (|S| - 1)P(\underline{U} \in \mathcal{U}_{\underline{z}}). \tag{28}$$

where $\underline{U}$ is a random vector which is uniform on $T_n(\delta,p_1)$ and

$$\mathcal{U}_{\underline{z}} \triangleq \{\underline{u} \in \mathcal{X}_1^n : (\underline{u},\underline{z}) \in J_n(\delta_1,q_1)\}.$$

Since $\underline{U}$ is uniform on $T_n(\delta,p_1)$ we have

$$P\{\underline{U} \in \mathcal{U}_{\underline{z}}\} = |\mathcal{U}_{\underline{z}}|\,|T_n(\delta,p_1)|^{-1}. \tag{29}$$

By a slight modification of Lemma 2.1.2 of [5] (reversing signs in expressions of the form $p(u) + \delta|\mathcal{U}|^{-1}$ and $p(u,v) + \delta|\mathcal{W}|^{-1}$), or by Lemma 2.1.6 of [11],

$$|\mathcal{U}_{\underline{z}}| \le \exp_2\{n[h_z(q_1) + f'(\delta_1)]\}. \tag{30}$$

where $f'(\delta_1) \to 0$ as $\delta_1 \to 0$ and $f'(\delta_1)$ is not a function of p. From property 2 of the δ-typical sets

$$|T_n(\delta,p_1)|^{-1} \le \exp_2[-n(\mathcal{H}(p_1) - c)], \tag{31}$$

so $$P\{\underline{U} \in \mathcal{U}_{\underline{z}}\} \le \exp_2[n(h_z(q_1) - \mathcal{H}(p_1) + f'(\delta_1) + c)]. \tag{32}$$

Substituting (32) into (28) we get

$$P[E_{\underline{z}}] < |S|\exp_2[n(h_z(q_1) - \mathcal{H}(p_1) + f'(\delta_1) + c)]$$

and using (27)

$$P[E_z] < \exp_2[n(h_z(q_1) - \bar{h}_z(p_1) + f'(\delta_1) + c - \gamma)] \tag{33}$$

$$\leq \exp_2[n(f'(\delta_1) + c - \gamma)] \tag{34}$$

where (34) follows from the definition of $\bar{h}_z(p_1)$. So $P[E_{\underline{z}}] \to 0$ as $n \to \infty$ if $\gamma > c + f'(\delta_1)$ for any $\underline{z} \in D$. The constants c and $f'(\delta_1)$ may be made arbitrarily small by choice of δ, hence any $R_1 > \bar{h}_z(p_1)$ is achievable and the rate region is as given by (22).

4.3 Converse to Coding Theorem

The converse to this result is exactly the same as for the result of Chapter 3. Only a single rate is possible for all $\pi \in \Lambda_1(p_1)$ and $R_1 \geq h_z(q_1)$ for each π (and its associated q_1) in the class by the converse to the side information problem hence $R_1 \geq \sup\{h_z(q_1) : \pi \in \Lambda_1(p_1)\} = \bar{h}_z(p_1)$.

CHAPTER 5

THE WYNER-ZIV PROBLEM

5.1 Introduction

Encoder 2 sends at rate $R_2 = \mathcal{H}(p_2)$, the entropy of source 2, so the decoder may be assumed to have $\{X_k^{(2)}\}$ exactly; therefore only the coding for source 1 need be considered. It is desired to determine the rate distortion function $R_1(D)$ for source 1; that is, the minimum rate R_1 such that the decoder can produce an r.v. $\hat{X}^{(1)}$ which satisfies $E\{d(X^{(1)},\hat{X}^{(1)})\} \leq D$. The distortion measure is assumed to be a finite single letter distortion measure on $\mathcal{X}_1 \times \hat{\mathcal{X}}_1$ which satisfies $d(u,u) = 0$ and $d(u,v) > 0$ for $u \neq v$ (see Fig. 4). Again an auxiliary random variable Z is required, so let q be the joint distribution of $X^{(1)}, X^{(2)}$, and Z and let q_i be the joint distribution of $X^{(i)}, Z$ induced by q.

If the joint distribution of $X^{(1)}$ and $X^{(2)}$ is known precisely then the rate distortion function is [4],

$$R^*(D) \triangleq \inf\{\mathcal{I}(q_1) - \mathcal{I}(q_2): \quad q \in Q(D)\} \tag{35}$$

where

$$Q(D) \triangleq \{q : X^{(2)} \to X^{(1)} \to Z \text{ and } \exists\, f : \mathcal{X}_2 \times \mathcal{Z} \to \hat{\mathcal{X}}_1 \ni E\{d(X^{(1)}, f(X^{(2)}, z))\} \leq D\} \tag{36}$$

and $\mathcal{I}(q_i)$ is the mutual information between $X^{(i)}$ and Z computed using the distribution q_i. Now any auxiliary random variable Z with joint distribution $q \in Q(D)$ may be described by a conditional distribution w where $w(u|v) = P\{Z = u | X^{(1)} = v\}$ since $q(x^{(1)},x^{(2)},z) = p(x^{(1)},x^{(2)})w(z|x^{(1)})$ by

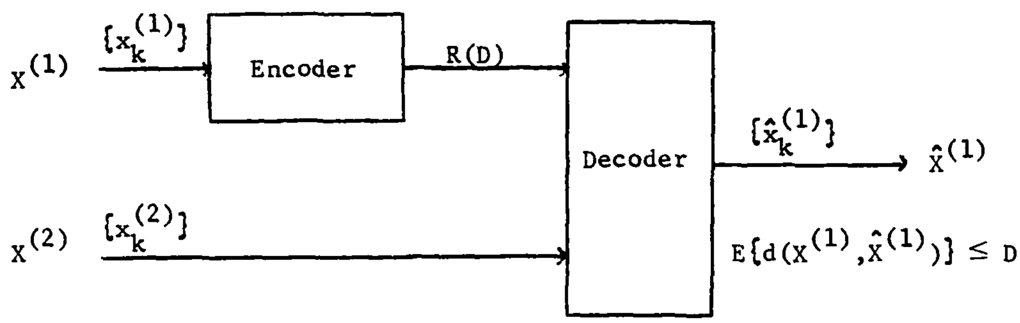


Figure 4. Configuration for the Wyner-Ziv problem.

the Markov property. So an alternative description of the rate distortion function is

$$R^*(D) = \inf\{\mathscr{I}(q_1) - \mathscr{I}(q_2) : w \in P(p,D)\} \tag{37}$$

where

$$P(p,D) = \{w : \sum_{v,z} F(v,z) \leq D\} \tag{38}$$

is defined in terms of the function F which is given by

$$F(v,z) = \min\{\sum_u w(z|u)p(u,v)d(u,\hat{u}) : \hat{u} \in \hat{\mathcal{X}}_1\} \tag{39}$$

These descriptions are equivalent for if $w \in P(p,D)$ then define the function $f(v,z) = \hat{u}^*$ where $\hat{u}^*$ is such that

$$\sum_u P\{X^{(1)} = u \mid X^{(2)} = v, Z = z\} d(u,\hat{u}^*) = \min\{\sum_u P\{X^{(1)} = u \mid X^{(2)} = v, Z = z\}$$
$$d(u,\hat{u}) : \hat{u} \in \hat{\mathcal{X}}_1\} \tag{40}$$

and this f will satisfy $E\{d(X^{(1)}, f(X^{(2)},Z))\} \leq D$ by the definition of $P(p,D)$. Conversely if $q \in Q(D)$ then the corresponding $w \in P(p,D)$ because f as defined by (39) minimizes the contribution of each (v,z) pair ($v \in \mathcal{X}_2$, $z \in \mathcal{Z}$) to the distortion, which minimizes the total distortion.

Now assume that the encoder knows only that $p \in \Lambda$ and define

$$W(D) \triangleq \bigcap_{\pi \in \Lambda_1(p_1)} P(\pi,D) \tag{41}$$

where

$$\Lambda_1(p_1) \stackrel{\Delta}{=} \{\pi \in \Lambda : \pi_1 = p_1\}. \tag{42}$$

Then the set

$$\bar{\mathcal{R}} = \{ (R_1,D) : R_1 \geq \inf\{ \sup\{\mathcal{J}(q_1) - \mathcal{J}(q_2) : \pi \in \Lambda_1(p_1)\} : w \in W(D)\} \tag{43}$$

is achievable. This is clear since by the definition of W(D) the decoder can find an r.v. $\hat{X}^{(1)}$ with $\{d(X^{(1)},\hat{X}^{(1)})\} \leq D$ once it has Z, and the random coding proof of the Wyner-Ziv result in [5] shows that any $R_1 \geq \mathcal{J}(q_1) - \mathcal{J}(q_2)$ allows Z to be decoded. For a given w, $R_1 \geq \sup\{\mathcal{J}(q_1) - \mathcal{J}(q_2) : \pi \in \Lambda(p_1)\}$ so Z may be recovered for any $\pi \in \Lambda_1(p_1)$ and $\bar{\mathcal{R}}$ is achievable. However, $\bar{\mathcal{R}}$ does not necessarily contain any point on any of the $R^*(D)$ curves for individual $\pi \in \Lambda_1(p_1)$. For any given w, the distribution π which achieves the maximum rate need not also have the worst distortion. So the above result does not establish the existance of robust codes.

The following example shows that in fact robust codes do not in general exist for the Wyner-Ziv problem. That is, in general it is not possible to construct a code for any class of sources which achieves a point (R,D) on the rate-distortion function $R^*(D)$ for one of the sources and distortion no greater than D for the other sources.

5.2 Counter Example for Robust Coding

Here $|\mathcal{X}_i| = 2$ $(i = 1,2)$. The class Λ is composed of two pairs of sources. One has distribution π given by $\pi(0,0) = \pi(1,1) = \frac{1}{2}(1-\beta_s)$ and $\pi(1,0) = \pi(0,1) = \frac{1}{2}\beta_s$. The other has distribution $\hat{\pi}$ which is given by

$\hat{\pi}(0,0) = \frac{1}{2}(1-\beta_z)$, $\hat{\pi}(0,1) = \frac{1}{2}\beta_z$, $\hat{\pi}(1,0) = 0$ and $\hat{\pi}(1,1) = \frac{1}{2}$. In these definitions β_s and β_z are constants in $[0,1]$. The distribution π corresponds to a doubly symmetric binary source (DSBS) and $\hat{\pi}$ is defined by a "Z-channel" between $X^{(1)}$ and $X^{(2)}$, a "Z-channel" being one with only a single possible crossover (Fig. 5). Note that $\pi_1 = \hat{\pi}_1$ so $\Lambda = \Lambda_1(\pi_1) = \Lambda_1(\hat{\pi}_1)$.

Two different conditional distributions of Z given $X^{(1)}$ are necessary and these are denoted by w and $\hat{w}$. The distributions π, $\hat{\pi}$, w, and $\hat{w}$ are used to define three joint distributions on $X^{(1)}$, $X^{(2)}$, and Z as below

$$q(u,v,z) \triangleq \pi(u,v)w(z|v) \tag{44a}$$

$$q'(u,v,z) \triangleq \hat{\pi}(u,v)w(z|v) \tag{44b}$$

$$q''(u,v,z) \triangleq \hat{\pi}(u,v)\hat{w}(z|v) \tag{44c}$$

for $u \in \mathcal{X}$, $v \in \mathcal{X}_2$, and $z \in \mathcal{Z}$. Joint distributions induced on $X^{(i)}$ and Z by these distributions are denoted by q_i, q_i', and q_i'' respectively. For the conditional distribution w and source π, if $q(u,v,z) = \pi(u,v)w(z|v)$ we may define

$$r(\pi,w) \triangleq \mathcal{I}(q_1) - \mathcal{I}(q_2) \tag{45}$$

and

$$\delta(\pi,w) \triangleq \sum_{v,z} \min\{\sum_u q(u,v,z)d(u,u) : \hat{u} \in \mathcal{X}_1\}. \tag{46}$$

So $r(\pi,w)$ and $\delta(\pi,w)$ are the rate and distortion achieved by a source π and conditional distribution w.

To achieve a particular point on the rate distortion curve $R^*(D)$ for π (the DSBS) the conditional distribution of Z given $X^{(1)}$ must correspond to a specific binary symmetric channel (BSC) with a fixed

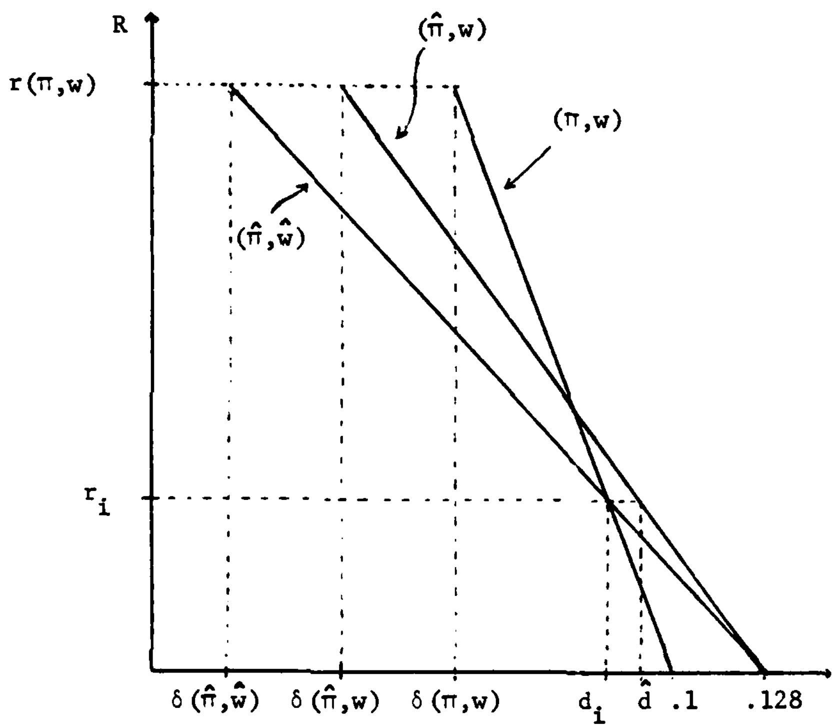


Figure 5. Time-sharing performance of three source-test channel pairs.

level of time sharing. Any other distribution will do strictly worse. Examination of the derivation of $R^*(D)$ for the DSBS in [4] makes this clear (see Appendix B).

Now fix $\beta_s = .1$ (which defines π) and solve for d^*, the distortion at which time sharing begins for source π. Solving equation (26) of [4] numerically, we find $d^* = .00752$ and $R^*_\pi(d^*) = 0.4238$. ($R^*_\pi(D)$ is the Wyner-Ziv rate distortion function for source π.) Then let the test channel w be a BSC with crossover probability d^* so that $\delta(\pi,w) = d^*$, the auxiliary r.v. Z must be defined by this test channel and some fixed level of time sharing. The distortion at rate equal to zero is $\beta_s = .1$ for π, so $R^*_\pi(D)$ for $D \geq d^*$ is as in Fig. 5. Next find a value β_s (which defines $\hat{\pi}$) such that

$$r(\hat{\pi},w) = r(\pi,w) \ . \tag{47}$$

The value of β_z satisfying this constraint is .256, and $\delta(\hat{\pi},w) = .00472$ is the distortion which results when test channel w is used with source $\hat{\pi}$. The distortion for $\hat{\pi}$ at rate equal to zero is $\frac{1}{2}\beta_z = .128$ so the time sharing performance of w and $\hat{\pi}$ is as shown in Fig. 5.

Next we wish to find a test channel $\hat{w}$ which yields lower distortion at the same rate for source $\hat{\pi}$. If we define $\hat{w}(1|0) = P\{Z = 1|X^{(1)} = 0\} = .01465$ and $\hat{w}(0|1) = .00465$ then $\delta(\hat{\pi},\hat{w}) = .0042 < \delta(\hat{\pi},w)$ and $r(\hat{\pi},\hat{w}) = r(\hat{\pi},w)$ as desired. Note that the pair $[r(\hat{\pi},\hat{w}),\delta(\hat{\pi},\hat{w})]$ is not necessarily on the rate distortion curve for $\hat{\pi}$, but it is an upper bound; that is,

$$R^*_{\hat{\pi}}[\delta(\hat{\pi},\hat{w})] \leq r(\hat{\pi},\hat{w}) \ . \tag{48}$$

Referring to Fig. 5, we let d_i denote the distortion and r_i denote the rate of the point corresponding to the intersection of the $(w - \pi)$ curve and the $(\hat{w} - \hat{\pi})$ curve. Test channel w applied to source $\hat{\pi}$ achieves distortion $\bar{\hat{d}} > d_i$ at rate r_i.

Let $\bar{D} \triangleq \max\{d(u,v) : u \in \mathcal{X}_1, v \in \hat{\mathcal{X}}_1\}$ and define a conditional distribution distance $d_s(w',w'')$ by

$$d_s(w',w'') \triangleq \sum_{u,v} |w'(u|v) - w''(u|v)| \ . \tag{49}$$

Pick $\epsilon > 0$ such that $\hat{d} - d_i > \epsilon\bar{D} > 0$. If a new auxiliary r.v. Z has w' as its conditional distribution given $X^{(1)}$ and w' is such that $d_s(w',w) > \epsilon$, then this r.v. will have distortion strictly greater than d_i (at rate r_i) when used for source π (the DSBS) by the proof in Appendix A. Yet if $d_s(w',w) \leq \epsilon$ then the distortion achieved for source $\hat{\pi}$ is

$$\sum_{x,z} \min\{\sum_u w'(z|u)\hat{\pi}(u,x)d(u,v) : v \in \hat{\mathcal{X}}_1\} \geq \hat{d} - \epsilon\bar{D} > d_i \ . \tag{50}$$

In either case the distortion for one of the two sources $\{\pi,\hat{\pi}\}$ exceeds d_i, hence no single r.v. can achieve distortion $\leq d_i$ for both sources at rate r_i. This does imply that no robust code exists, as can be seen from the converse to the Wyner-Ziv result (Section III of [4]). If a sequence of codes of increasing block length n exists which achieves asymptotically a distortion d_i at rate r_i for both π and $\hat{\pi}$ then eq. (55) of [4] holds and so here exists a sequence of random variables $\{Z_j^{(n)} : 1 \leq j \leq n\}$ which satisfy $Z_j^{(n)} \to X_j^{(1)} \to X_j^{(2)}$ and achieve distortions $\Delta_j^{(n)}$ such that

$$\lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} \Delta_j^{(n)} = d_i \ . \tag{51}$$

Let $w_j^{(n)}$ be the conditional distribution corresponding to $Z_j^{(n)}$. By the proof in Appendix B, the point (r_i,d_i) is achieved for source π by a unique conditional distribution w and time-sharing parameter $\theta \in [0,1]$. Hence the sequence of conditional distributions $w_j^{(n)}$ must approach this conditional distribution w with a particular level of time sharing. But this possibility has been eliminated, because w with time sharing achieves a distortion $\hat{d} > d_i$ for source $\hat{\pi}$. Hence no robust code exists.

5.3 Special Case where Robust Coding is Possible

If the problem is restricted to sets Λ with the property that there is one worst source π^* such that $\pi^*(u,v) = \sum_y \pi(u,y)\, c_\pi(v|y)$ for all $\pi \in \Lambda$, where c_π is a conditional distribution, then a robust coding technique similar to that of the side information case may be obtained. The coding method in this case is simply to code for the worst source π^*. Let $\hat{q}$ be a joint distribution of $X^{(1)}$, $X^{(2)}$, and Z such that the distribution of $X^{(1)}$ and $X^{(2)}$ induced by $\hat{q}$ is π^*. Then the (R,D) pairs achieved are given by

$$R(D) \geq \inf\{\mathcal{I}(\hat{q}_1) - \mathcal{I}(\hat{q}_2) \quad w \in W(D)\} \tag{52}$$

where $\hat{q}(u,v,z) = \pi^*(u,v)w(z|u)$ and

$$W(D) \triangleq \{w : \sum_{v,z} \min[\sum_u \hat{q}(u,v,z)d(u,\hat{u}) : \hat{u} \in \mathcal{X}_1] \leq D\} \;. \tag{53}$$

Now for any $\pi \in \Lambda$ there exists a c_π such that $\sum_y \pi(u,y)c(v|y) = \pi^*(u,v)$ so $\mathcal{I}(q_2) \geq \mathcal{I}(\hat{q}_2)$ and $\mathcal{I}(\hat{q}_1) - \mathcal{I}(\hat{q}_2) \geq \mathcal{I}(q_1) - \mathcal{I}(q_2)$ for fixed w (note that $\mathcal{I}(\hat{q}_1) = \mathcal{I}(q_1)$). The r.v. Z defined by w may be recovered if $R \geq \mathcal{I}(q_1) - \mathcal{I}(q_2)$ so the decoder may recover Z regardless which $\pi \in \Lambda$ is the true distribution.

The decoder can derive a function f: $\mathcal{X}_2 \times \mathcal{Z} \to \mathcal{X}_1$ such that $E\{d(X^{(1)},f(X^{(2)},Z))\} \le D$ as follows. Given π^* and π, a c_π can be derived satisfying $\sum_y \pi(u,y)c_\pi(v|y) = \pi^*(u,v)$. Then, using c_π an r.v. $\hat{X}^{(2)}$ may be derived from $X^{(2)}$ such that the joint distribution of $\hat{X}^{(2)}$ and $X^{(1)}$ is π^*. Then the decoding function for source π^* will yield the desired result. Since source π is better than π^* in some sense, this technique may not yield the minimum distortion, and the actual distortion achieved may be less than that of the worst source.

The performance of the robust coding can be given for the class $\Lambda = \{\mathrm{DSBS}(\theta):\theta \in [\theta_1,\theta_2]\}$ with Hamming distortion measure $0<\theta_1 \le \theta_2 < \frac{1}{2}$. The source DSBS($\theta_2$) is the worst source π^*, and its rate distortion function may be achieved by the coding technique of part IIa of [4]. If the source in effect is π^* then the $R^*_{\pi^*}(D)$ curve is achieved. This is the solid line in Fig. 6. For the sources DSBS(θ) with $\theta < \theta_2$ there are two cases. If $\theta > d^*$, the distortion at which time sharing begins, then the performance is the same as for DSBS(θ_2) up to d^*, and for $D \ge d^*$ the performance is better, as the distortion at rate zero is θ (dotted curve in Fig. 6). The maximum distortion for any source DSBS(θ) is θ, so if $\theta < d^*$ then the performance is given by the dashed curve in Fig. 6. Similar performance is achieved in the case of the doubly symmetric M-ary source, where tight bounds on $R^*(D)$ are known [6], as this is also a totally ordered set.

5.4 <u>Note on Evaluation of the Wyner-Ziv Rate Region</u>

One of the difficulties in coding for the Wyner-Ziv problem is that explicit formulas for the rate distortion function are known only for doubly symmetric sources. Even numerical evaluation of the Wyner-Ziv

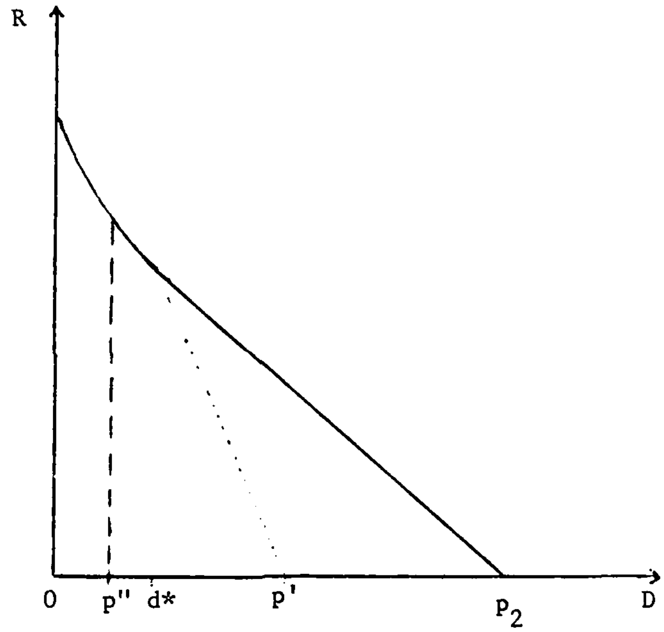


Figure 6. Performance of robust coding for the DSBS.

rate region is difficult as the constraint

$$\sum_{v,z} \min\{\sum_{u} \pi(u,v)w(z|u)d(u,\hat{u});\ \hat{u} \in \hat{\mathcal{X}}_1\} \le D \tag{54}$$

is not in general convex. For example suppose the source is a DSBS(.1) and consider conditional distributions w_1, w_2, w_3 defined by $w_1(0|0)$ $w_1(0|0) = w_2(1|1) = .5$, $w_1(1|) = w_2(0|0) = 1$, and

$$w_3(u|v) \triangleq \tfrac{1}{2}[w_1(u|v) + w_2(u|v)] \ .$$

Then the distortion achieved by w_1 and w_2 is .075, but the distortion achieved by w_3 is .10. So the set of conditional distributions w which satisfy the constraint (54) with D = .075 is not a convex set.

CHAPTER 6

ROBUST CODING FOR THE SLEPIAN-WOLF PROBLEM

6.1 Statement of Problem

The problem here is the same as that of Chapter 3 except that encoder 2 is not constrained to use $R_2 > \mathcal{H}(p_2)$. The outputs of both sources must still be decoded with arbitrarily low probability of error. As before the joint distribution p is in a class Λ, and the i-th encoder knows the i-th marginal p_i, and the decoder knows p. The set $\Lambda_i(p_i)$ is defined by

$$\Lambda_i(p_i) = \{\pi \in \Lambda: \pi_i = p_i\}$$

as before, and we define

$$\Lambda(p) \triangleq \{\pi \in \Lambda: \pi_i = p_i;\ i = 1,2\}. \tag{55}$$

6.2 An Outer Bound on the Achievable Rate Region

For a given marginal any code must have a fixed rate. So for each subset of sources $\Lambda(p)$ there will be a single rate pair (R_1,R_2). By the converse to the source coding theorem $R_1 + R_2 \geq \mathcal{H}(\pi)$ for every $\pi \in \Lambda$. So if the true distribution for the source is p then $R_1 + R_2 \geq \sup\{\mathcal{H}(\pi): \pi \in \Lambda(p)\}$. Defining

$$\bar{h}_i(p_i) \triangleq \sup\{h_i(\pi): \pi \in \Lambda_i(p_i)\}$$

we also know $R_i \geq \bar{h}_i(p_i)$ where $h_1(\pi) \triangleq \mathcal{H}(\pi) - \mathcal{H}(\pi_2)$ and $h_2(\pi) \triangleq \mathcal{H}(\pi) - \mathcal{H}(\pi_1)$, by the robust coding result for the corner points of the Slepian-Wolf region. So an outer bound on the achievable rate region is

$$\begin{aligned} \bar{\mathcal{R}} \triangleq \{(R_1,R_2):\ & R_1 + R_2 \geq \sup[\mathcal{H}(\pi): \pi \in \Lambda(p)], \\ & R_i \geq \sup[h_i(\pi): \pi \in \Lambda_i(p_i)],\ i = 1,2\}. \end{aligned} \tag{56}$$

Now neither encoder can determine the set $\Lambda(p)$, so the i-th encoder cannot determine its rate R_i from (56). However, since the decoder knows the joint distribution p, any subset of $\bar{\mathcal{R}}$ for which the i-th encoder may determine R_i is achievable by the random coding argument used in Chapter 3. Note that the set $\bar{\mathcal{R}}$ is a function of p, so we actually have a family of sets $\bar{\mathcal{R}}(p)$ which bound the achievable rate regions for each source $p \in \Lambda$.

6.3 Some Sets of Achievable Rates

Three different coding techniques which yield sets of achievable rates are considered in this section. In the first technique $R_1 = k$ and $R_2 \geq \sup\{\mathcal{H}(\pi) - k : \pi \in \Lambda_2(p_2)\}$, where k is a constant chosen beforehand. If the source distribution is π we must have $R_1 \geq h_1(\pi)$ so k must satisfy

$$k \geq \sup\{h_1(\pi) : \pi \in \Lambda\} \tag{57}$$

Also $R_2 \geq \bar{h}_2(p_2)$ so if we define

$$\mathcal{R}_{k1} \stackrel{\Delta}{=} \{(R_1,R_2) | R_1 \geq k, R_2 \geq \max(\bar{h}_2(p_2), \sup[\mathcal{H}(\pi)-k : \pi \in \Lambda_2(p_2)])\} \tag{58}$$

where k satisfies (57), then $\mathcal{R}_{k1}$ is a subset of $\bar{\mathcal{R}}$ and R_i is a function only of p_i so $\mathcal{R}_{k1}$ is achievable. Reversing the roles of the encoders yields another set $\mathcal{R}_{k2}$.

Another set of achievable rates is given by time sharing between the corner points of the Slepian-Wolf region derived in Chapter 3. Let $\beta \in [0,1]$ be the time sharing parameter selected beforehand and define

$$\mathcal{R}_\beta \stackrel{\Delta}{=} \{(R_1,R_2) : R_1 \geq \sup[R_1(\beta,\pi) : \pi \in \Lambda_1(p_1)],$$

$$R_2 \geq \sup[R_2(1-\beta,\pi) : \pi \in \Lambda_2(p_2)]\} \tag{59}$$

where

$$R_i(\beta,\pi) \stackrel{\Delta}{=} \beta\mathcal{H}(\pi_i) + (1-\beta)h_i(\pi). \tag{60}$$

The i-th encoder can clearly determine R_i so we need only show $\mathcal{R}_\beta \subset \tilde{\mathcal{R}}$. Note $\Lambda(p) \subset \Lambda_i(p_i)$ so if $(R_1,R_2) \in \mathcal{R}_\beta$ then

$$R_i \geq \sup\{R_i(\beta,\pi) : \pi \in \Lambda(p)\}$$

and

$$R_1 + R_2 \geq \sup\{\mathcal{H}(\pi) : \pi \in \Lambda(p)\}.$$

Also since $0 \leq \beta \leq 1$

$$R_1 \geq \sup\{h_i(\pi) : \pi \in \Lambda_i(p_i)\}$$

and so $\mathcal{R}_\beta \subset \mathcal{R}$.

A third technique is to choose an $\alpha > 0$ (α will represent the ratio R_2/R_1), and define

$$\beta(\pi) = \begin{cases} 0 & \alpha \leq \dfrac{h_2(\pi)}{\mathcal{H}(\pi_1)} \\[2ex] 1 & \alpha \geq \dfrac{\mathcal{H}(\pi_2)}{h_1(\pi)} \\[2ex] \dfrac{\alpha\mathcal{H}(p_1) - h_2(\pi)}{\alpha\mathcal{H}(\pi_1) - h_2(\pi) + \mathcal{H}(\pi_2) - \alpha h_1(\pi)} & \text{otherwise} \end{cases} \tag{61}$$

Then

$$\mathcal{R}_\alpha \triangleq \{(R_1,R_2) : R_1 \geq \sup[R_1(\beta(\pi),\pi) : \pi \in \Lambda_1(p_1)]$$

$$R_2 \geq \sup[R_2(1-\beta(\pi),\pi) : \pi \in \Lambda_2(p_2)]\} \tag{62}$$

is achievable using the same reasoning as for $\mathcal{R}_\beta$. Regions derived from these two techniques (for all possible α and β) are sketched in Figure 7.

The sets $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$ may be improved in the following way. If one of the encoders can determine from its marginal that the other encoder is using a rate higher than necessary it may reduce its own rate. Applied to $\mathcal{R}_\beta$ this yields the set

$$\mathcal{R}_{\beta 1} \triangleq \{(R_1,R_2) : R_1 \geq \sup[R_1(\beta,\pi) - \sup\{R_2(1-\beta,\hat{\pi}) : \hat{\pi} \in \Lambda_2(\pi_2)\}$$

$$+ R_2(1-\beta,\pi) : \pi \in \Lambda_1(p_1)],$$

$$R_2 \geq \sup[R_2(1-\beta,\pi) : \pi \in \Lambda_2(p_2)]\} \tag{63}$$

Since $\pi \in \Lambda_2(\pi_2)$ it is clear that $\mathcal{R}_{\beta 1}$ is at least as good as $\mathcal{R}_\beta$. To show $\mathcal{R}_{\beta 1} \subset \bar{\mathcal{R}}$ suppose that the source distribution is p. Then

$$R_1 \geq R_1(\beta,p) - \sup\{R_2(1-\beta,\pi) : \pi \in \Lambda_1(p_1)\} + R_2(1-\beta,p)$$

and

$$R_2 \geq \sup\{R_2(1-\beta,\pi) : \pi \in \Lambda_1(p_1)\}$$

so $R_1 + R_2 \geq \mathcal{H}(p)$ and $\mathcal{R}_{\beta 1} \subset \bar{\mathcal{R}}$. The same modification applied to $\mathcal{R}_\alpha$ yields

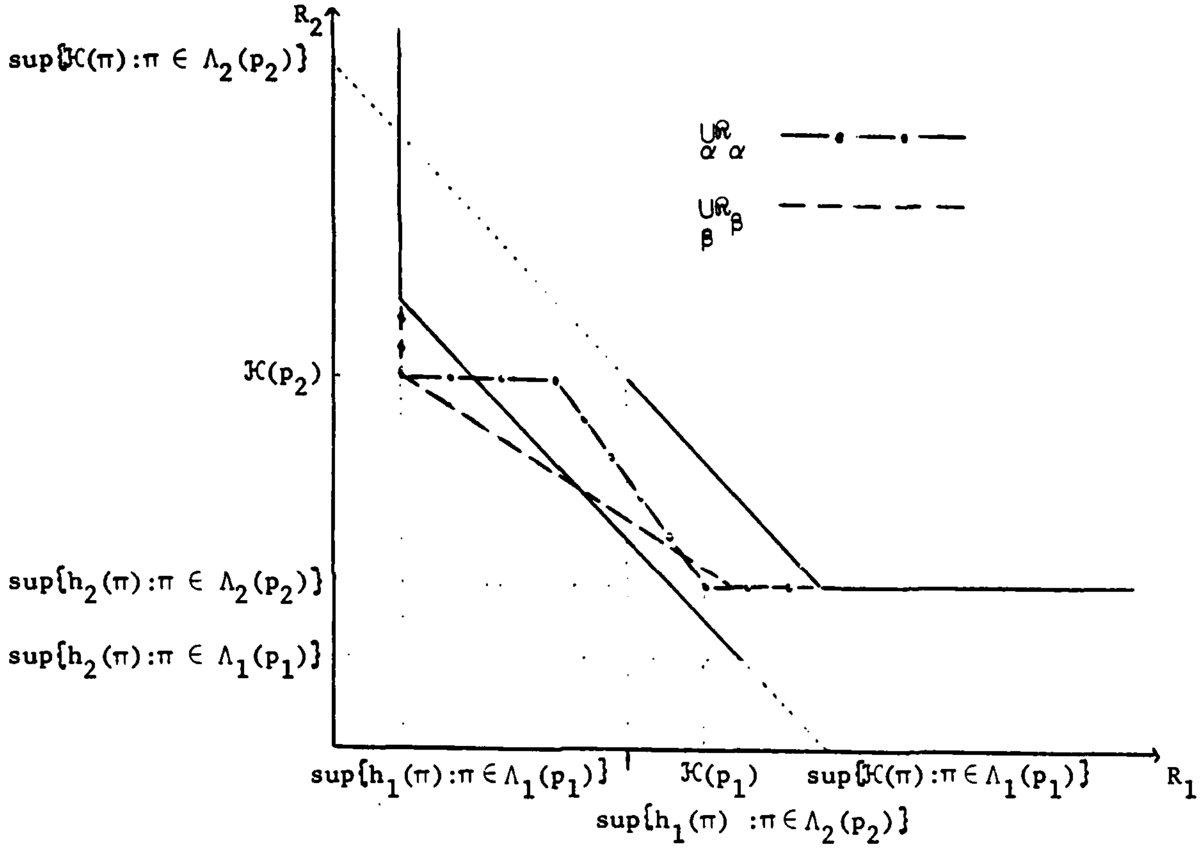


Figure 7. Sketch of regions $\mathfrak{R}_\alpha$ and $\mathfrak{R}_\beta$.

a set $\mathcal{R}_{\alpha 1}$ which is defined as $\mathcal{R}_{\beta 1}$ with β replaced by $\beta(p)$, and the same reasoning shows that it is achievable. Similar sets $\mathcal{R}_{\alpha 2}$ and $\mathcal{R}_{\beta 2}$ may be defined switching the roles of the encoders.

An example where $\mathcal{R}_{\alpha 1} \neq \mathcal{R}_{\alpha}$ and $\mathcal{R}_{\beta 1} \neq \mathcal{R}_{\beta}$ is given in Figure 8. Here Λ consists of two sources π and $\hat{\pi}$ such that $\pi_2 = \hat{\pi}_2$ but $\pi_1 \neq \hat{\pi}_1$. In Figure 8, π is the true distribution of the source. For this example $\mathcal{R}_{\alpha 2} = \mathcal{R}_{\alpha}$ and $\mathcal{R}_{\beta 2} = \mathcal{R}_{\beta}$ so the imporved sets are not uniformly better, nor are they necessarily equal. The Slepian-Wolf region for π is denoted $\mathcal{R}^*(\pi)$ in the figure.

The following example shows that even the union of all of these sets is not the achievable rate region. Here $\Lambda = \{\pi, \pi', \pi''\}$ and these sources have Slepian-Wolf rate regions as in Figure 9. Note that $\pi_1 = \pi_1'' \neq \pi_1'$, $\pi_2 = \pi_2' \neq \pi_2''$, $\mathcal{H}(\pi_1) < \mathcal{H}(\pi_1')$, and $\mathcal{H}(\pi_2) < \mathcal{H}(\pi_2'')$. The various sets are given in Figure 9 assuming that π is in effect. A better set $\hat{\mathcal{R}}$ may be derived as follows.

If π_i is the observed marginal then the i-th encoder codes as for $\mathcal{R}_{\alpha}$ but using π in place of $\Lambda_i(\pi_i)$. So if π is the actual source then the set $\hat{\mathcal{R}}$ (see Figure 9) is achieved. If π' is in effect encoder 1 can determine this and use rate

$$R_1 = R_1(\beta(\pi'), \pi') + R_2(\beta(\pi'), \pi') - R_2(\beta(\pi), \pi)$$

and the π' rate region is achieved. If π'' is in effect the encoder 2 does similarly and the π'' rate region is achieved. So even the union of these sets of achievable rate pairs is not the achievable rate region.

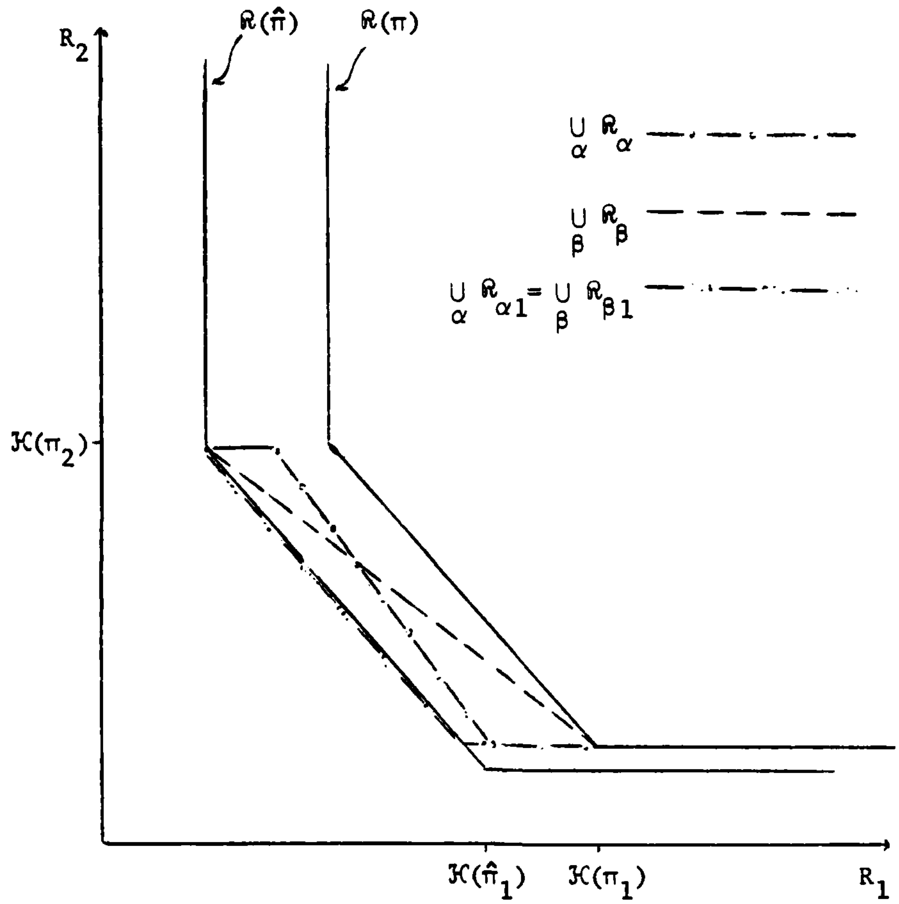


Figure 8. Example where $\mathcal{R}_{\alpha 1}$ and $\mathcal{R}_{\beta 1}$ are strictly better than $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$.

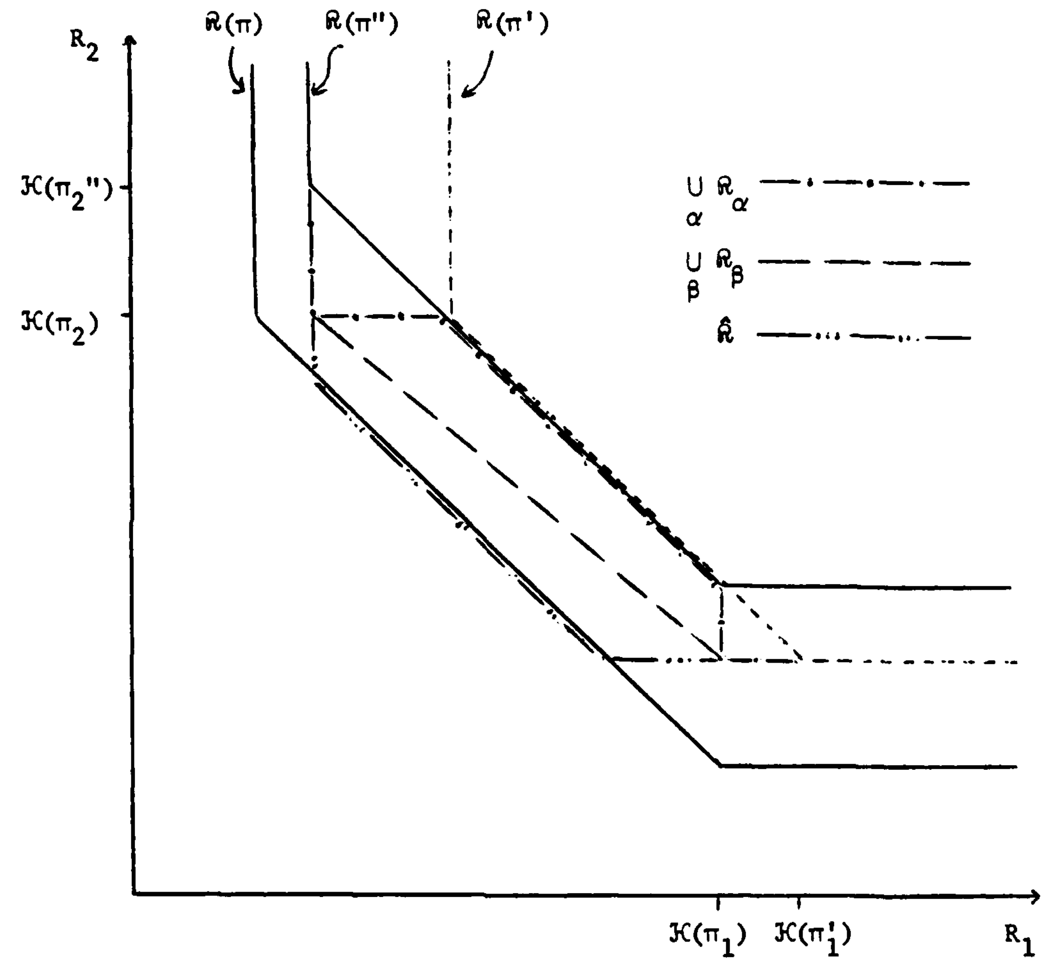


Figure 9. Example showing that the union of these sets of achievable rates is not the entire rate region. [Note here $\mathcal{R}_\alpha = \mathcal{R}_{\alpha i}$ and $\mathcal{R}_\beta = \mathcal{R}_{\beta i}$, $i = 1,2$, $\mathcal{R}_{k1} = \mathcal{R}(\pi')$ and $\mathcal{R}_{k2} = \mathcal{R}(\pi'')$.]

CHAPTER 7

CONCLUSION

We have considered a number of problems in source coding for pairs of correlated discrete memoryless sources for the situation in which the distribution of the source outputs is not precisely known by the encoders or the decoder. We have assumed only that the source is in some class. For problems where decoding with arbitrarily low distortion is required we showed that universal coding (as defined for single-user problems) is not generally possible. We defined robust codes as codes which achieve the the optimum performance for one source in the class and achieve performance which is no worse (i.e., no larger rate or distortion) for the other sources in the class. For the side information problem and the corner point of the Slepian-Wolf region the rate for one of the encoders does not depend on the class of possible joint distributions. In these two problems we found that the optimum performance was achieved by robust coding for a particular subset of the class of sources. A bound was given for the optimum performance for the Slepian-Wolf problem, but the achievable rate region was not determined.

For each of the above problems the converse to the robust coding theorem was provided by the converse to the corresponding problem in which the source distribution is known. That is, the converse is derived from the fact that the converse for some individual source in the class (or in some subset of the class as determined by the encoders) requires the rate to be at least that of the positive coding theorem for the entire class. This technique may not be used to prove the converse for the Wyner-Ziv problem because the counterexample of Section 5.2 shows that for some classes of sources the optimal performance is strictly worse than that for all individual sources in the class.

APPENDIX A

MODIFICATIONS NECESSARY FOR THE PROOFS OF CHAPTERS 3 AND 4 WITHOUT THE ASSUMPTION OF KNOWN MARGINALS

Here we assume that the i-th encoder has no apriori knowledge of the i-th marginal p_i and that the decoder has no apriori information on the joint distribution p. If f and g are real-valued functions defined on a finite set $\mathcal{X}$ then we define

$$d_m(f,g) = \max\{|f(u) - g(u)| : u \in \mathcal{X}\} \ . \tag{65}$$

Let $\mathcal{M}$ represent the space of all discrete memoryless sources with finite alphabet $\mathcal{X}_1 \times \mathcal{X}_2$. If we define

$$\mathcal{J} \triangleq \{\pi \in \mathcal{M} : \pi(u,v) \geq \epsilon' \quad \text{for all } u,v \in \mathcal{X}_1 \times \mathcal{X}_2\} \tag{66}$$

then for any $p \in \mathcal{M}$

$$\inf\{d_m(p,\pi) : \pi \in \mathcal{J}\} < 2\epsilon' \ . \tag{67}$$

For some initial time (i.e., based on an initial block of source output symbols) the i-th source estimates the marginal p_i. Call this estimate π_i. This initial time may be chosen such that

$$P\{|p_i(u) - \pi_i(u)| > \epsilon'' \text{ for some } u \in \mathcal{X}_i\} < \delta' \ . \tag{68}$$

For any ϵ'', $\delta' > 0$. From (67) a source $\hat{p} \in \mathcal{J}$ exists such that $d_m(\pi_i, \hat{p}_i) < 2|\mathcal{X}_i|\epsilon'$ and so for this $\hat{p}$ we have

$$P\{|p_i(u) - \hat{p}_i(u)| > \epsilon \text{ for some } u \in \mathcal{X}_i\} < \delta' \tag{69}$$

by the triangle inequality where $\epsilon \overset{\Delta}{=} \epsilon'' + 2|\mathcal{X}|\epsilon'$ and $|\mathcal{X}| = \max\{|\mathcal{X}_i|;\ i=1,2\}$. The marginal $\hat{p}_i$ is then used to encode some block of source outputs. Assuming that the entire block is in error if $\hat{p}_i$ is not within ϵ of p_i for all $u \in \mathcal{X}_i$, this introduces a probability of error less than $2\delta'$ for the coding scheme.

The decoder makes an estimate π of the joint distribution p by using rates $R_i = \mathcal{H}(\hat{p}_i)$, $(i=1,2)$, for some initial time. The decoder then finds an $\hat{p} \in \mathscr{J}$ which is within $2\epsilon'$ of π. This initial time may be chosen such that

$$P\{|p(u,v) - \hat{p}(u,v)| > \bar{\epsilon} \text{ for some } (u,v) \in \mathcal{X}_1 \times \mathcal{X}_2\} < \delta' \tag{70}$$

where $\bar{\epsilon} \overset{\Delta}{=} \epsilon|\mathcal{X}|^{-1}$ and δ' and ϵ are defined in (69). As with the marginals we now assume $d_m(p,\hat{p}) \le \bar{\epsilon}$ adding a block probability of error of δ'.

The encoders may obtain sets of sequences with properties similar to those of the $T_n(\delta,p_i)$ using these estimated marginal distributions as follows. Let the i-th encoder define $\underline{u}$ to be <u>typical</u> if

$$|n^{-1} N(u|\underline{u}) - \hat{p}_i(u)| < \delta|\mathcal{X}_i|^{-1} + \epsilon \tag{71}$$

for all $u \in \mathcal{X}_i$ where $\hat{p}_i$ satisfies $|\hat{p}_i(u) - p_i(u)| < \epsilon$. Call the set of such sequences $T_n'(\delta,\hat{p}_i)$. Then from (1) and (71)

$$T_n(\delta,p_i) \subset T_n'(\delta,\hat{p}_i) \tag{72}$$

hence

$$P[\underline{u} \in T_n'(\delta,\hat{p}_i)] \to 1 \quad \text{as } n \to \infty \tag{73}$$

for any $\epsilon,\delta > 0$. Also

$$T_n'(\delta,\hat{p}_i) \subset T_n(\delta + 2\epsilon|\mathcal{X}|,p_i) \tag{74}$$

so

$$\exp_2[n(\mathcal{H}(p_i) - k)] < |T_n'(\delta,p_i)| < \exp_2[n(\mathcal{H}(p_i) + k)] \tag{75}$$

where

$$k \triangleq -[\delta + 2\epsilon|\mathcal{X}|]\log \epsilon' \tag{76}$$

$$\geq -[\delta|\mathcal{X}_i|^{-1} + 2\epsilon]\sum_u \log \hat{p}_i(u) \tag{77}$$

Note that k may be made arbitrarily small by choice of ϵ, δ, and ϵ'. Here we use the lower bound on $\hat{p}_i(u)$ provided by (66).

The decoder defines the set $J_n'(\delta,\hat{p})$ of jointly typical pairs $(\underline{u},\underline{v})$ as those satisfying

$$|n^{-1} N(u,v|\underline{u},\underline{v}) - p(u,v)| < \delta|\mathcal{X}_1|^{-1}|\mathcal{X}_2|^{-1} + \bar{\epsilon} \tag{78}$$

and so

$$\exp_2[n(\mathcal{H}(p) - k)] < J_n'(\delta,\hat{p}) < \exp[n(\mathcal{H}(p) + k)] \tag{79}$$

since from (76)

$$k \geq -[\delta|\mathcal{X}_1|^{-1}|\mathcal{X}_2|^{-1} + \epsilon|\mathcal{X}|^{-1}] \sum_{u,v} \log p(u,v) \tag{80}$$

Finally define the set

$$\Lambda_i'(\hat{p}_i) \triangleq \{\pi \in \Lambda : \max_u |\pi_i(u) - \hat{p}_i(u)| < \epsilon\}. \tag{81}$$

In coding for the corner point of the Slepian-Wolf region (Chap. 3) these sets $T_n'(\delta,\hat{p}_i)$ and $\Lambda_i'(\hat{p}_i)$ may be used and yield the same results in the limit as $\epsilon \to 0$ (recall Λ is assumed closed). Note that all information theoretic quantities are continuous bounded functions of the probability distributions (here the alphabets are finite) so the use of the estimate $\hat{p}$ instead of p to compute them will result in a term added to the required rates of the form $f(\epsilon)$ (where $f(\epsilon) \to 0$ as $\epsilon \to 0$).

The proofs of the three Lemmas used in Chapter 4 may be easily modified to hold for the typical sets of this section simply by using (73) and (75) in place of properties 1 and 2 of Chapter 2. The functions of δ in (24) and (30) become functions of ϵ and δ which approach zero as ϵ and δ approach zero. These functions remain independent of $\hat{p}$ because of the uniform bound of (66).

APPENDIX B

PROOF OF THE COUNTER EXAMPLE FOR THE WYNER-ZIV PROBLEM

Here we show that the auxiliary r.v. Z in the Wyner-Ziv problem must be defined as the output of a BSC with X as input in order for Z to be optimal for the DSBS. The notation used and equations referred to in this appendix are from [4].

Referring to equations (42) and (43), if the d_z are not identical then we have

$$R_x = I(X;Z) - I(Y;Z) \geq \theta \sum_{z \in A} \lambda_z G(\delta_z) > \theta G\Big(\sum_{z \in A} \lambda_z d_z\Big)$$

since $G(d) \triangleq h(p_0 * d) - h(d)$ is a strictly convex function of d for fixed $p_0 \in (0,\frac{1}{2})$. By definition,

$$d_z = E[d(X,\hat{X})|Z = z] = P\{X \neq \Upsilon(Z)|Z = z\}$$

and $\Upsilon(z)$ has range $\{0,1\}$. If the d_z are identical (for all $z \in A$) then by the definition of $\Upsilon(z)$, A may be divided into two sets $A_i = \{z:\Upsilon(z) = i\}$, $i = 0,1$, such that all $z \in A$ are equivalent in the sense that

$$P\{X = x, Y = y|Z = z_0\} = P\{X = x, Y = y|Z = z_1\} \text{ for all } x,\ y,$$

if z_0 and $z_1 \in A_i$. So for the $z \in A$ the channel defining Z is a BSC with parameter d_z. The set A^c may be regarded as time sharing since from (36)

$$E[d(X,\hat{X})|Z \in A^c] = p_0,$$

and distortion p_0 may be achieved with $R_x = 0$. Furthermore, the last line of (40) is exactly the rate required by time sharing between the BSC with output $\Upsilon(z)$ and parameter d_z and no channel at all. For equality to hold in (40), so that the rate for Z is the same as this time sharing, we must have

$$\sum_{z \in A^c} [H(Y|Z = z) - H(X|Z = z)] \; P\{Z = z\} = 0$$

which implies

$$P[Z = z|X = 0] = P[Z = z|X = 1]$$

for $z \in A^c$. So for a r.v. Z to achieve the lowest possible rate, no information about X can be given by the event $Z \in A^c$. So the optimal channel must be a combination of a BSC which is used with some probability θ, and a channel which gives no information which is used with probability $(1 - \theta)$.

We have shown that a BSC with some level of time sharing must be used to attain optimal performance. It is easily seen that the parameter of the BSC and level of time sharing are unique. Up to the point where time sharing begins no two BSC's will achieve the same rate, since G(d) is strictly decreasing. G(d) is strictly convex so in the time sharing region no other BSC with time sharing will do as well as the BSC with crossover probability d^* by the definition of d^*.

REFERENCES

1. D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Transactions on Information Theory, vol. IT-19, pp. 471-480, 1973.

2. A. Wyner, "On source coding with side information at the decoder," IEEE Transactions on Information Theory, vol. IT-21, pp. 294-300, 1975.

3. R. Ahlswede and J. Körner, "Source coding with side information and a converse for the degraded broadcast channel," IEEE Transactions on Information Theory, vol. IT-21, pp. 629-637, 1975.

4. A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Transactions on Information Theory, vol. IT-22, pp. 1-10, 1976.

5. T. Berger, "Multiterminal source coding," in The Information Theory Approach to Communications, G. Longo, ed., CISM Courses and Lectures #299, Springer-Verlag, Wien, New York, 1978.

6. K. Mackenthun, Jr., "An evaluation of the Wyner-Ziv rate region," Proceedings of the 17th Annual Allerton Conference on Communications, Control, and Computing, pp. 274-283, 1979.

7. J. C. Kieffer, "Some universal multiterminal source coding theorem," submitted for publication.

8. T. Berger, Rate Distortion Theory: A Mathematical Basis for Data Compression, Prentice-Hall, Englewood Cliffs, N. J., 1971.

9. R. E. Blahut, "Computation of channel capacity and rate distortion functions," IEEE Transactions on Information Theory, vol. IT-18, pp. 460-473, 1972.

10. R. J. McEliece, The Theory of Information and Coding, Addison-Wesley, Reading, Mass., 1977.

11. J. Wolfowitz, Coding Theorems of Information Theory, 3rd Ed., Springer-Verlag, Berlin, New York, 1978.

12. L. D. Davisson, "Universal source coding," IEEE Transactions on Information Theory, vol. IT-19, pp. 783-795, 1973.

13. D. Neuhoff, R. Gray, and L. Davisson, "Fixed rate universal source coding with a fidelity criterion," IEEE Transactions on Information Theory, vol. IT-21, pp. 511-523, 1975.

14. D. Sakrison, "Worse sources and robust codes for difference distortion measures," IEEE Transactions on Information Theory, vol. IT-21, pp. 301-309, 1975.

LME

-8